Research Report

On Multidimensional Item Response Theory: A Coordinate-Free Approach

Tamás Antal

Research & Development

July 2007
RR-07-30

On Multidimensional Item Response Theory: A Coordinate-Free Approach

Tamás Antal

ETS, Princeton, NJ

July 2007

Abstract

A coordinate-free definition of complex-structure multidimensional item response theory (MIRT) for dichotomously scored items is presented. The point of view taken emphasizes the possibilities and subtleties of understanding MIRT as a multidimensional extension of the classical unidimensional item response theory models. The main theorem of the paper is that every monotonic MIRT model looks the same; they are all trivial extensions of univariate item response theory.

Key words: MIRT, geometric methods, compensatory models

Acknowledgments

The author would like to extend his gratitude to Paul Holland for his the insights and constant encouragement. The author is also indebted to Henry Chen for fruitful discussions. This paper was inspired by a thought-provoking presentation by Mark Reckase at ETS in September 2006. The author is indebted for the lively discussion during this presentation. Andreas Oranje and Shelby Haberman provided useful reviews of the manuscript.

1 Introduction

Complex-structure multidimensional item response theory (MIRT) is built on the idea that a single item, however simple it might be, carries the possibility of an inner structure. In usual terminology one speculates that it is possible to measure several cognitive areas with one item. The number of cognitive areas so measured may vary among items, even though usual models assume that it is fixed for a collection of items (a test) and let a factor analysis type procedure decide on the number and mixture of cognitive areas measurable by the items.

The point of view taken in this note is that any unidimensional item response theory (IRT) model can be thought of as a specialization of an MIRT model. Hence, the major task is to identify how much of the well-established tools and nomenclature of unidimensional IRT can be preserved in the multidimensional context and, from the other direction, how different multidimensional notions may specialize to the same unidimensional entity. When the latter happens, that is when two different multidimensional objects yield the same unidimensional specialization, then both multidimensional notions could be considered proper generalizations of the underlying unidimensional quantity. A careful study should then be devised to decide which generalization is more appropriate with respect to the application at hand.

There is, on the other hand, the possibility of not finding proper multidimensional generalization for some unidimensional notions. This topic also deserves careful research and understanding.

This paper only considers what is termed *complex-structure* MIRT. Usually, IRT models have two components: the item likelihood and the population distribution. In *simple-structure* MIRT, one item represents only one dimension and without a multivariate population distribution the entire likelihood of the model would factor as a product of univariate pieces. In complex-structure MIRT, this factorization is impossible, by definition, irrespective of population model chosen.

The structure of the paper is as follows: a short overview of unidimensional IRT is followed by the *absolute*, that is, coordinate-free definition of MIRT. The connection with the usual approach is also shown via a discussion of two widely accepted models. Then the development of the main thesis follows. This paper proves that MIRT models are all alike and they all can be obtained as a trivial extension of an appropriate unidimensional IRT model. Two sections on some thoughts about capturing cognitive dimensions and on understanding the role of the notion of dimension-wise independence close the presentation.

2 Unidimensional Item Response Theory

To make the generalization to the multidimensional framework easier, some features of unidimensional IRT are summarized first. Measurement takes place during the formation of the response matrix $X \in M_{N\times I}(\mathbb{N})$ with elements $x_{ni} \in \mathbb{N}$ for student $n = 1, \dots, N$ and item $i = 1, \dots, I$. In a dichotomous setting (which is assumed throughout the paper to simplify the presentation), $x_{ni} = 1$ if student n responded correctly to item i, otherwise it is zero. As a major simplification of the modeling of the cognitive process, it is assumed that students response to an item is stochastically determined by students ability θ_n and item parameters $\beta_i := (a_i, b_i, c_i)$ via the item response function (Birnbaum, 1968).

$$P^{3\mathrm{pl}}(\theta_n, \beta_i) := \mathrm{Prob}(x_{ni} = 1 \mid \theta_n, \beta_i) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta_n - b_i)}}. \tag{1}$$

There are, of course, many different item response functions in use: the three-parameter logistic (3PL) model is chosen here only as an illustration. The other substantial simplification used in building the model is the assumption of independence of conditional probabilities $P^{3\mathrm{pl}}_{ni}$ across an arbitrary subset $S \subset \{1, \dots, N\} \times \{1, \dots, I\}$ of student-item pairs.

The two most popular models built out of these blocks are the joint unidimensional IRT and the marginal unidimensional IRT. Joint IRT states that the total likelihood depends explicitly on the ability of the given students:

$$L^{\mathrm{joint}}(X; \Theta, \mathcal{B}) = \prod_{n,i} P^{3\mathrm{pl}}(\theta_n, \beta_i)^{x_{ni}}(1 - P^{3\mathrm{pl}}(\theta_n, \beta_i))^{1-x_{ni}} \tag{2}$$

with corresponding log-likelihood:

$$\mathcal{L}^{\mathrm{joint}}(X; \Theta, \mathcal{B}) = \sum_{n,i} x_{ni} \log(P^{3\mathrm{pl}}(\theta_n, \beta_i)) + (1 - x_{ni}) \log(1 - P^{3\mathrm{pl}}(\theta_n, \beta_i)). \tag{3}$$

Here, Θ and $\mathcal{B}$ are the collections of all abilities and item parameters, respectively.
In marginal theory, the likelihood depends only on the distributional properties of student's population:

$$L^{\mathrm{marg}}(X; \mathcal{B}, \Phi) = \prod_n \int_{\mathbb{R}} \prod_i P^{3\mathrm{pl}}(\theta, \beta_i)^{x_{ni}}(1 - P^{3\mathrm{pl}}(\theta, \beta_i))^{1-x_{ni}} \mathrm{d}\mu_n(\theta) \tag{4}$$

with log-likelihood

$$\mathcal{L}^{\mathrm{marg}}(X; \mathcal{B}, \Phi) = \sum_n \log \int_{\mathbb{R}} \prod_i P^{3\mathrm{pl}}(\theta, \beta_i)^{x_{ni}}(1 - P^{3\mathrm{pl}}(\theta, \beta_i))^{1-x_{ni}} \mathrm{d}\mu_n(\theta), \tag{5}$$

where μ_n is the density measure of student n over $\mathbb{R}$ and Φ is the collection of distributional parameters for student's ability. In a parametric setting, usually μ_n is given as $d\mu_n(\theta) = \varphi_n(\theta)\mathrm{d}\theta$ with some distribution function φ_n.

The quantities

$$L_n^{\mathrm{st}}(X; \theta, \mathcal{B}) = \prod_i P^{3\mathrm{pl}}(\theta, \beta_i)^{x_{ni}}(1 - P^{3\mathrm{pl}}(\theta, \beta_i))^{1-x_{ni}} \tag{6}$$

and

$$L_i^{\mathrm{it}}(X; \Theta, \beta) = \prod_n P^{3\mathrm{pl}}(\theta_n, \beta)^{x_{ni}}(1 - P^{3\mathrm{pl}}(\theta_n, \beta))^{1-x_{ni}} \tag{7}$$

are the student and item likelihoods, respectively.

It is worthwhile to analyze the shape of the student likelihood function. It is a product of the conditional probabilities of the actual responses over all the items administered to the student. As a function of θ, the probability of the correct response increases when the actual response is correct and decreases for an incorrect actual response. As a consequence, a student likelihood will increase if all the actual responses are correct and decrease if all the actual responses are incorrect. This in turn pushes the location of the maximum likelihood solution for the given student to plus or minus infinity. For the item likelihood, a similar statement holds. For a generic response pattern, the student likelihood can be well-approximated by the density function of the normal distribution (especially when the number of items is large enough) and its curvature will be inversely proportional to the asymptotic standard error of the ability estimates.

3 Multidimensional Item Response Theory

3.1 Basic Models

Even though widely investigated, MIRT is not yet widespread as an operational model. Hence, identifying the major players among the competing MIRT models is difficult. Here, only two models are discussed, one by Whitely (1980) and another one by Reckase (1997). With an item, a vector of discriminations $a \in \mathbb{R}^D$ and a vector of difficulties $b \in \mathbb{R}^D$ are associated. With these the functional representation of the *dimension-wise independent* MIRT model of Whitely has the form

$$f^w_{a,b} : \mathbb{R}^D \to [0, 1], \quad \theta \mapsto f^w_{a,b}(\theta) = \prod_{d=1}^{D} \frac{1}{1 + e^{-a_d(\theta_d - b_d)}}. \tag{8}$$

If the conditional probability of passing the dth dimension of the item is given by $\frac{1}{1+e^{-a_d(\theta_d-b_d)}}$, then (8) can indeed be understood as the joint probability of passing *all* the independent dimensions of the item. Unless there are separate observed scores for each dimension, language such as *correct response on dimension d* cannot be used. To address this lack, this paper uses the phrase *passing a dimension term*, which may refer to an unobservable event.

McKinley and Reckase (1982) (see also Reckase, 1997) put forward a model that takes the functional representation

$$f^{sp}_{a,b} : \mathbb{R}^D \to [0,1], \quad \theta \mapsto f^{sp}_{a,b}(\theta) = \frac{1}{1+e^{-\langle a \mid \theta \rangle - b}}, \tag{9}$$

where a is as before and $b \in \mathbb{R}$. $\langle x \mid y \rangle = \sum_{d=1}^{D} x_d y_d$ is the usual scalar product of $x, y \in \mathbb{R}^D$. This paper uses the term *scalar product MIRT* to refer to this model. See Figure 1.

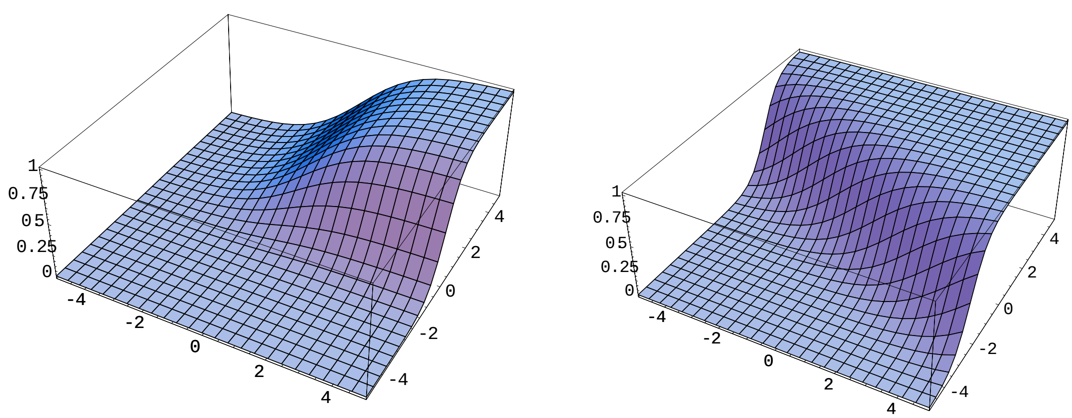

***Figure* 1 Dimension-wise independent and scalar product MIRT hypersurface.**

3.2 Definition of Multidimensional Item Response Theory

The goal in this section is to give a definition of MIRT with as few assumption as possible. MIRT postulates that with a single item, multiple cognitive abilities could be detected. To accommodate this idea, one has to change the model for the ability space from the one-dimensional vector space $\mathbb{R}$ to a finite-dimensional vector space V_θ. While any finite-dimensional vector space V is linearly isomorphic to $\mathbb{R}^D$ for $D = \dim(V)$ [see (10) for an explicit way of constructing such an isomorphism], this isomorphism is not canonical (there is no unique isomorphism $V \to \mathbb{R}^D$). Because of this and other reasons that will become clear, $\mathbb{R}^D$ is not used here as a mathematical model of ability space.

The reader unfamiliar with these notions is referred to Halmos (1974) for an excellent introduction to linear algebra. Also, an intuitive understanding of the basic notions of smooth manifolds should help understanding of what follows, although it not strictly necessary. Among the many fine references to the topic, the interested reader may find Warner (1971) useful.

The basic object in unidimensional IRT is the item response function (IRF) and its graph, the item response curve (IRC). The graph of a function $f : A \to B$ is a subset $\mathrm{graph}(f) = \{(x, f(x)) \in A \times B \mid x \in A\}$ of $A \times B$. IRC is a one-dimensional smooth submanifold of $\mathbb{R} \times [0,1]$. While there is a scaling freedom even in the one-dimensional case (e.g., the [in]famous 1.7 multiplier in the logistic models), the possibility of ambiguous interpretation is minimal and one may use the functional (IRF) and the geometrical (IRC) representation almost interchangeably.

In the multidimensional case, however, the matter is not so straightforward. As this paper shows, the functional representation and the geometric representations are different in a subtle way. One way to keep the presentation coordinate-free in multidimensional IRT is to postulate that the theory is given by an *item response hypersurface (IRHS)*. As in the unidimensional case, the IRHS is used to express the probability of correct response given an ability in V_θ.

Before this paper defined this notion, some notations should be fixed. For any $v \in V_\theta$, the *ray* of v is defined to be the line $\mathbb{R} \cdot v$ in V_θ determined by v: $\mathbb{R} \cdot v = \{\lambda v \in V_\theta \mid \lambda \in \mathbb{R}\}$. Similarly, for $v, w \in V_\theta$, the v-directed line going through w is defined by

$$w + \mathbb{R} \cdot v = \{w + \lambda v \in V_\theta \mid \lambda \in \mathbb{R}\}.$$

For the notion of IRHS, there is then the following:

Definition 1 *A dichotomous item response hypersurface (IRHS) is a* $D = \dim(V_\theta)$ *dimensional smooth submanifold* M *of* $V_\theta \times [0,1]$, *so that for any two vectors* $v, w \in V_\theta$ *the intersection of* $(w + \mathbb{R} \cdot v) \times [0,1]$ *and* M *is a graph of a monotonic function* $w + \mathbb{R} \cdot v \to [0,1]$.

A MIRT model *is given when an IRHS is given.*

Note, that while $w + \mathbb{R} \cdot v$ is not canonically isomorphic to $\mathbb{R}$, monotonicity of the map

$$f_{v,w} : w + \mathbb{R} \cdot v \to [0,1], \quad \lambda \mapsto f_{v,w}(w + \lambda v)$$

can be unambiguously defined by requiring that either $f_{v,w}(w + \lambda v) \leq f_{v,w}(w + \mu v)$ or

$f_{v,w}(w + \lambda v) \geq f_{v,w}(w + \mu v)$ for all $\lambda, \mu \in \mathbb{R}$ whenever $\lambda \leq \mu$. The notation $f_v = f_{v,0}$ is used here. Figure 2 shows the intersection of $(w + \mathbb{R} \cdot v) \times [0, 1]$ and M in two dimensions.

To understand the definition better, first assume that v is arbitrary and $w = 0$ in the definition above. Then, the line $w + \mathbb{R} \cdot v = \mathbb{R} \cdot v$ can be understood as an ability direction. The monotonicity requirement of Definition 1 asks for the natural feature that, as the ability given by v increases, the probability of the correct response increases as well. For nonzero w, the requirement is equivalent to the conditional probability of correct response being monotonic with respect to one ability when the rest of the abilities are fixed to a certain, not necessarily zero value. To be precise, one should say that for $w \neq 0$ there exists a basis of V_θ so that the monotonicity requirement reads as the interpretation above. Furthermore, for any basis of V_θ, Definition 1 will ensure the monotonicity of the conditional probability of correct response for any ability direction given any fixed values for the rest of the ability directions (as defined by the basis).

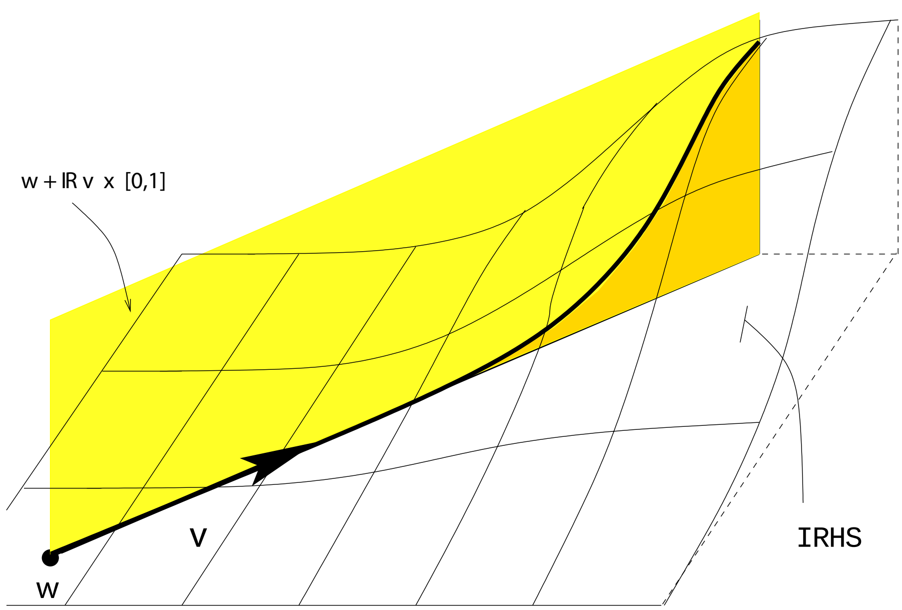

***Figure* 2 Intersection of $(w + \mathbb{R} \cdot v) \times [0, 1]$ and IRHS. The intersection is the bold curve, which is required to be monotonic.**

Note also that the collection of maps $f_{v,w}$ for $v, w \in V_\theta$ defines the IRHS completely. For this reason, the notation f^M, or f if no confusion may arise, is used for the function describing the IRHS $M \subset V_\theta \times [0, 1]$.

One may be tempted to object to the use of notions like manifold and hypersurfaces. It is very important to note, however, that the conditional probability of correct response has been given by a hypersurface in the usual MIRT literature as well. One major difference in terminology is that it was still called *surface* in any dimension, which is a correct usage only in Dimension 2. In higher dimensions, the object at hand is a hypersurface, a special case of higher dimensional manifolds.

A basis $\mathfrak{v} = (v_1, \ldots, v_D)$ in V_θ defines a unique isomorphism

$$i_\mathfrak{v} : V_\theta \to \mathbb{R}^D, \quad \sum_{i=1}^{D} \lambda_i v_i \mapsto \sum_{i=1}^{D} \lambda_i e_i, \qquad (\lambda_i \in \mathbb{R}), \tag{10}$$

where $(e_i)_{i=1}^D$ is the standard basis of $\mathbb{R}^D$: $(e_i)_j = \delta_{ij}$ (δ_{ij} is the Kronecker delta). This isomorphism can be trivially extended to a diffeomorphism

$$\psi_\mathfrak{v} : V_\theta \times [0,1] \to \mathbb{R}^D \times [0,1], \quad (v,t) \mapsto (i_\mathfrak{v}(v), t) \tag{11}$$

and via this diffeomorphism the IRHS may be transferred from $V_\theta \times [0,1]$ to $\mathbb{R}^D \times [0,1]$. Now, in $\mathbb{R}^D \times [0,1]$ the image of the IRHS may be given by the graph of a smooth function $f : \mathbb{R}^D \to [0,1]$. Note, however, the important difference between using a functional representation like this latter one and using the hypersurface representation directly in $V_\theta \times [0,1]$. The functional representation depends on the basis chosen to establish the diffeomorphism $\psi_\mathfrak{v}$ and different bases may result in different functional representations.

It is tempting to extend this definition to polytomous multidimensional items by defining the *polytomous collection of item response hypersurfaces* for a polytomous item by requiring that the above discussed intersection be a collection of unidimensional polytomous item response curves as produced by some unidimensional polytomous IRT model (e.g., Muraki's [1992) partial credit model). The investigation of this possibility is postponed for a forthcoming paper.

3.3 Properties of Item Response Hypersurfaces

This section proves the main theorem of the paper. For the sake of transparency, this section starts with the two-dimensional case, which is then followed by the more involved general theory.

3.3.1 Two-Dimensional Case

Using the monotonicity of the model, it is possible to prove an interesting elementary property.

Lemma 1 *In any two-dimensional MIRT model, there exists a line in V_θ through the origin so that f_v is constant.*

Proof: Choose a vector, $v \in V_\theta$. Note that if $\inf_{\lambda\in\mathbb{R}} f_v(\lambda v) = \sup_{\lambda\in\mathbb{R}} f_v(\lambda v)$, the lemma is proved and the sought after line is $\mathbb{R} \cdot v$. Therefore, one may assume that $\inf_{\lambda\in\mathbb{R}} f_v(\lambda v) < \sup_{\lambda\in\mathbb{R}} f_v(\lambda v)$. For such a vector, either $\lim_{\lambda\to\infty} f_v(\lambda v) = \sup_{\lambda\in\mathbb{R}} f_v(\lambda v)$ or $\lim_{\lambda\to-\infty} f_v(\lambda v) = \sup_{\lambda\in\mathbb{R}} f_v(\lambda v)$. Let P be the set of vectors satisfying the first and N be the set of vectors satisfying the second condition. Both of these sets are nonempty, and by continuity, both sets are open. Also, they are clearly disjointed. Therefore, there is a vector $u \in V_\theta$ so that $u \notin N \cup P$. Along the line $\mathbb{R} \cdot u$ the function f is constant. ■

Note that the proof only uses monotonicity with $w = 0$. Utilizing it for general w, the same argument provides the following:

Lemma 2 *In any two-dimensional MIRT model, through* any *point $w \in V_\theta$ there exists $v \in V_\theta$ so that, along the v-directed line going through w, the function $f_{v,w}$ is constant.*

The term *w-constant line*, or simply *constant line*, is introduced here for the v-directed line going through w as in Lemma 2.

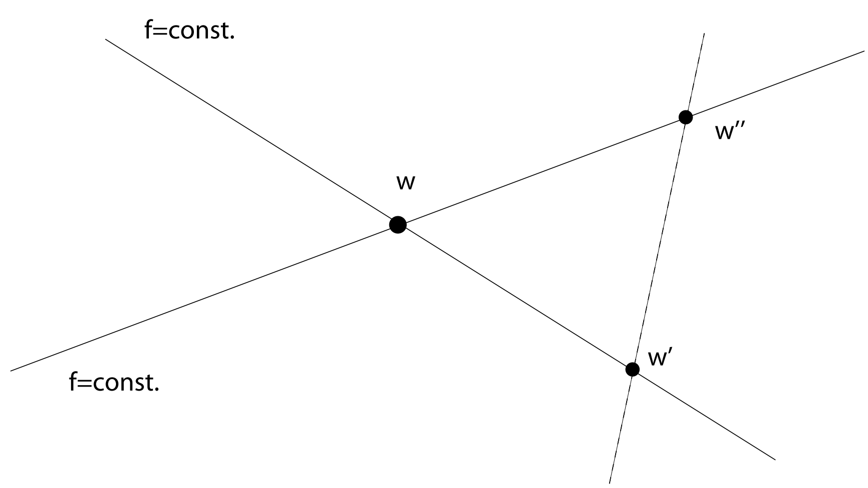

***Figure* 3 Nonunique constant line results in constant MIRT model.**

Analyzing the properties of these constant lines further, one can see that they are actually parallel to one another, which leads to the following:

Lemma 3 *Let $w, w' \in V_\theta$ be two points. Let $v, v' \in V_\theta$ be the corresponding directions of the two constant lines. Then $v = \mu v'$ for some $\mu \in \mathbb{R}$.*

Proof: First, note that if there is a point $w \in V_\theta$ so that there exist two w-constant lines, then the model is trivial (f is constant) and the statement is true. For, let w' and w'' be the intersections of a general position line in V_θ with the two w-constant lines, respectively. Because f is monotonic along this line, $f(w') = f(w'')$, f is constant between w' and w'', that is, $f(tw' + (1-t)w'') = f(w')$ for all $t \in [0, 1]$. Using this argument for every line in a general position proves that f is constant everywhere (see Figure 3).

Now assume that the constant lines through w and w' are unique. If the two lines are not parallel, then they will have an intersection, and an argument similar to the previous one shows that f is constant. ■

The corollary of the previous observation is given in Theorem 1.

Theorem 1 *Any two-dimensional MIRT model is a trivial extension of a unidimensional IRT model.*

Proof: Lemma 3 showed that a two-dimensional IRHS is nothing but a collection of parallel lines. Let $v \in V_\theta$ be the direction of these lines. By choosing a transversal $\mathbb{R} \cdot u$ (a line that intersects all of them) to this collection, the IRHS can be given by the function $f_u : \mathbb{R} \cdot u \to [0, 1]$. One can express an arbitrary $w \in V_\theta$ as a unique linear combination $w = \mu u + \lambda v$ and write

$$f(w) = f(\mu u + \lambda v) = f_u(\mu u). \tag{12}$$

This function f_u can be thought of as a unidimensional IRT model. ■

3.3.2 D-Dimensional Case

Technically, the D-dimensional case is not that much more complicated than the two-dimensional one. It is just much more difficult to visualize the corresponding geometric objects. As pointed out earlier, the conditional probability surface is not two-dimensional, so strictly speaking it is not a surface in higher dimensions. Three-dimensional training does not allow one to “see” objects in higher dimensions. The formalism built in the previous section, however, is applicable, with appropriate modifications, to this situation as well.

The proof of Lemma 1 works for any dimensions. Applying the monotonicity argument for arbitrary (v, w) as above proves the corresponding

Lemma 4 *In any MIRT model there exists a hyperplane H_w in V_θ through $w \in V_\theta$ so that f_{H_w} is constant.*

Proof: Here, f_{H_w} is the restriction of f to the hyperplane H_w. As before, $w = 0$ is proven explicitly; the general case follows the same argument. As before, the open sets P and N are denoted, with $P = -N$. Exclude the trivial case of $P = \emptyset$. It is clear that $P \cup N \neq V_\theta$. Locally the boundary of P (the closure of P minus P) is a $D-1$ dimensional submanifold ($D = \dim V_\theta$). Therefore there exists a collection $(c_1, \ldots, c_{D-1})$ of points in $V_\theta \backslash (P \cup N)$ so that $(c_1 - w, \ldots, c_{D-1} - w)$ spans a hyperplane H_w. Now, along the line segment joining any point on $\mathbb{R} \cdot (c_i - w)$ with any point on $\mathbb{R} \cdot (c_j - w)$ for some $i \neq j$, the restriction of f should be constant (Figure 3). Repeating this argument for each pair of line segments shows that along the entire hyperplane f is constant. If there is another hyperplane with this property, then $P = \emptyset$, which is excluded. ■

Now, a $D-1$ dimensional hyperplane is to a D-dimensional space as a line is to the plane. Using this intuition, it is not difficult to adapt the formal proof of Lemma 3 to prove Lemma 5:

Lemma 5 *Let $w, w' \in V_\theta$ be two points. Let $H_w, H_{w'} \subset V_\theta$ be the corresponding two constant hyperplanes. Then H_w and $H_{w'}$ are parallel.*

Now one can rephrase the main theorem in arbitrary dimension.

Theorem 2 *Any MIRT model is a trivial extension of a unidimensional IRT model.*

For the sake of explicitness, write f^M for an arbitrary IRHS in terms of univariate IRT model. A transversal $u \in V_\theta$ is fixed to the collection of constant hyperplanes. First, observe that for any $w \in V_\theta$ there is a unique decomposition $w = \mu u + \lambda v$ with $v \in H_w$. Then

$$f^M(\mu u + \lambda v) = f^M_u(\mu u). \tag{13}$$

Note that if the usual 2PL or 3PL models are chosen, the construction yields the scalar product model. It is also interesting to note that the MIRT generalization of the Rasch model is equivalent to the generalization of the 2PL model. This is because of the following: while within the univariate Rasch model, one may assume that the slope is fixed, but when more dimensions are considered simultaneously, the assumption of equal slopes is not valid. The relative positions of slopes to one another should be determined during the estimation procedure in lack of a priori information.

This kind of models were called generalized compensatory models (GMIRT) in Zhang and Stout (1999). The link function of an IRHS as GMIRT is f_u^M.

3.3.3 Absolute Functional Representation for the Scalar Product Model

A notable feature of the scalar product model is that using the dual of a vector space it can be defined without referring to coordinates even in its functional form. First, recall that the *dual* V^* of a finite-dimensional vector space V is the finite-dimensional vector space of the same dimension of linear maps $V \to \mathbb{R}$:

$$V^* := \{p : V \to \mathbb{R} \mid p \text{ is linear}\}. \tag{14}$$

The *duality* is the obvious map

$$(\ \mid\ ) : V^* \times V \to \mathbb{R},\ (p, v) \mapsto (\ p \mid v) := p(v). \tag{15}$$

That is, for any $p \in V^*$ and $v \in V$ the quantity $(\ p \mid v)$ is a real number. It is important to note that the duality, unlike a scalar product, does not involve any choice.

Now, if in MIRT the choice is made that the ability is modeled by the vector space V_θ as before and the item is modeled by the discrimination $a \in V_\theta^*$ in the dual space and a real number b, then the IRHS of the model is given as the graph of the following function:

$$f_{a,b}^d : V_\theta \to [0, 1],\ f_{a,b}^d(\theta) := \frac{1}{1 + e^{-(a \mid \theta) - b}}. \tag{16}$$

In addition to its very satisfying and elegant nature, this model has the computational advantage of having the same functional representation in *any* coordinate system. As is shown later, the dimension-wise independent model does not share this nice invariance property.

3.3.4 Interpretation of the Main Theorem

The statement of the main theorem excludes many existing MIRT models from the pool of monotonic MIRT models. The author's reading of the main theorem is that the only relevant MIRT model is the one defined in (13). This interpretation is backed by the fact the useful estimation methods exist only for the scalar product model, the most relevant of the above extensions (Reckase, 1997). In the view of Theorem 2, there seems to be a good reason behind that. It seems that lack of monotonicity prevents one from maximizing the likelihood function of MIRT models excluded by this approach. This certainly defines a valid future research direction. Also, the existence of an elegant coordinate-free functional representation makes the scalar product model even more appealing.

On the other hand, model building always has many steps that cannot be entirely backed by theoretical considerations. The process sometimes is dictated by personal preferences and tastes. It is possible that some readers may not be willing to accept the requirement of monotonicity as formulated in Definition 1 as a crucial and necessary feature of an MIRT model. For those readers the main theorem is interpreted a bit differently. First, note the close connection between the notion of *compensatory* model to monotonicity. Usual terminology is that the model is compensatory if the probability of the correct response may be high even with the lack of ability in all but one dimension. That is, sufficiently high ability in one dimension is able to compensate for the lack of it in other dimensions. In fact, compensatory property follows from monotonicity as an easy application of Theorem 2. If compensatory property is understood in a sense that it is true in any coordinate system, then the reverse is also true, and the two notions are equivalent. With this in mind the theorem states that *any compensatory MIRT model is a direct generalization of a univariate IRT model.*

In either way, Theorem 2 establishes a prominent role for the scalar product model as an MIRT model.

3.4 Estimation in Multidimensional Item Response Theory

The paper now restricts attention to the scalar product model. A typical student likelihood (with $D = 2$) is given as the right-hand side graph in Figure 4. As in unidimensional IRT, the maximum place of this function plays a special role in the estimation of MIRT model parameters. A curious feature of this graph is that a pronounced imbalance can be observed between the standard errors of the two ability estimates. Here, standard error is understood as the inverse of the curvature of the graph at the maximum place. There is a well-identified direction in which the standard error is minimal, and in the direction orthogonal to this, the standard error appears to be much bigger. One may even say that, despite these efforts, the model shows definite signs of unidimensionality. The reason behind this is very simple. A student likelihood is formed as a product of probabilities of the actual responses given by item response hypersurfaces similar to the one shown on the RHS of Figure 1. These hypersurfaces are always increasing towards the first quadrant (correct response) or towards the third quadrant (incorrect response). Hence, the product of these will be the ridge in Figure 4. It is a ridge because the observed response is either correct or incorrect, and no distinction is made between events of the students using only one of the dimensions correctly during the assessment. In other words, since there is no observed data for the different dimensions, the model will not be able to provide two distinct, meaningful estimates for the abilities of the person on the different cognitive dimensions.

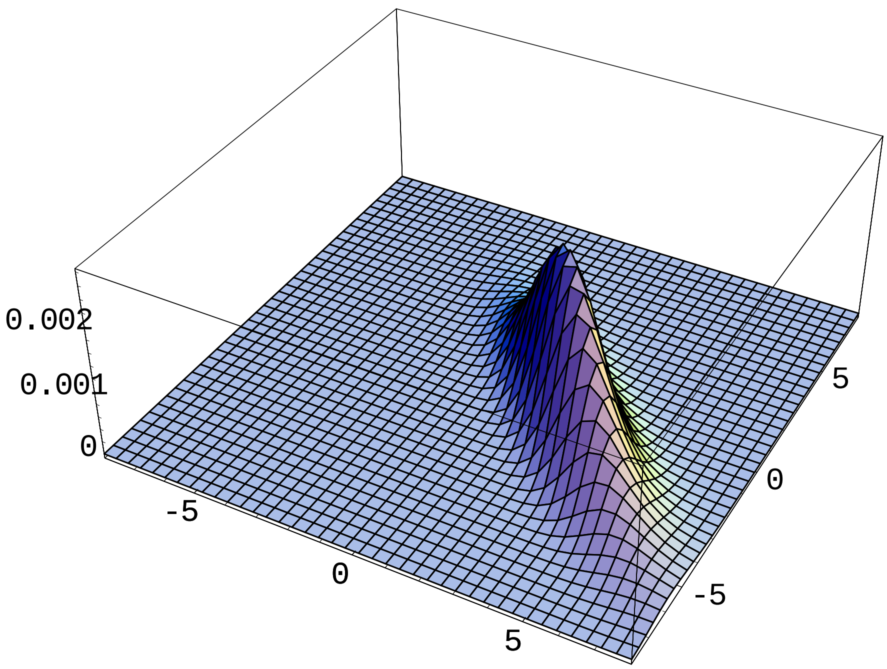

***Figure* 4 Scalar product MIRT student likelihood.**

3.5 Dimension-Wise Independence

The careful reader should have noticed that, concerning one particular point, the presentation is not faithful to its own principles. That is, the notion of dimension-wise independence was used without any discussion of its invariance or coordinate-system independence. It is easy to see that the dimension-wise independent model does not satisfy the requirement of monotonicity, therefore it would not be considered it as a valid MIRT model. On the other hand, it might be useful to see explicitly how badly the the functional representation of the dimension-wise independent model behaves and so appreciate the niceties of the scalar product model even more.

Invariance of dimension-wise independence for the the model

$$f^w_{a,b} : \mathbb{R}^D \to [0,1], \quad f^w_{a,b}(\theta) = \prod_{d=1}^{D} \frac{1}{1+e^{-a_d(\theta_d-b_d)}} \tag{17}$$

would mean that the factorization property holds in any other coordinate systems.

Mathematically, this would require that, for any invertible matrix $G \in \mathrm{GL}(D)$ (G expresses change of coordinates), is a function

$$h^G : \mathbb{R}\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}, \quad (a_1, b_1, t) \mapsto h^G(a_1, b_1, t)$$

with a pair of invertible matrices $U, V \in \mathrm{GL}(D)$, so that when $\theta = G \cdot \theta'$ ($\theta, \theta' \in \mathbb{R}^D$), there is have a factorization

$$f^w_{a,b}(\theta) = \prod_{d=1}^{D} h^G((Ua)_d, (Vb)_d, \theta'_d), \tag{18}$$

that is,

$$\begin{aligned} f^w_{a,b}(\theta) &= f^w_{a,b}(G\cdot\theta') \\ &= \prod_{d=1}^{D} \frac{1}{1+e^{-a_d(\sum_{d'=1}^{D} g_{dd'}\theta'_{d'}-b_d)}} \\ &= \prod_{d=1}^{D} h^G(a'_d, b'_d, \theta'_d), \end{aligned} \tag{19}$$

with $a'_d = (Ua)_d$ and $b'_d = (Vb)_d$. The role of U and V is to ensure that the function h^G is the same for all factors in the product by allowing this function to depend on different linear combinations of the elements of a and of b.

To show that this is too much to ask for in general, first assume that a factorization $f(x,y) = h(x)g(y)$ holds for some function f so that $h(0) \neq 0$ and $g(0) \neq 0$. Then,

$$\begin{aligned} h(x) &= \frac{f(x,0)}{g(0)}, \\ g(y) &= \frac{f(0,y)}{h(0)}, \end{aligned}$$

and

$$\frac{1}{h(0)g(0)} = \frac{f(x,y)}{f(x,0)f(0,y)}. \tag{20}$$

Now, for the sake of concreteness, take $D = 2$ and $a = (a_1, a_1) \in \mathbb{R}^2$ and $b = (0, 0) \in \mathbb{R}^2$. Also, take $G = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. With these, (19) becomes

$$\frac{1}{1 + e^{-a_1(\theta'_1+\theta'_2)}} \cdot \frac{1}{1 + e^{-a_1(\theta'_1-\theta'_2)}} = h(\theta'_1)g(\theta'_2) \tag{21}$$

with some $h, g : \mathbb{R} \to \mathbb{R}$. From (20) the function

$$\frac{1}{h(0)g(0)} = \frac{\frac{1}{1+e^{-a_1(\theta'_1+\theta'_2)}} \cdot \frac{1}{1+e^{-a_1(\theta'_1-\theta'_2)}}}{\frac{1}{(1+e^{-a_1\theta'_1})^2} \cdot \frac{1}{1+e^{-a_1\theta'_2}} \cdot \frac{1}{1+e^{+a_1\theta'_2}}} \tag{22}$$

should be constant. This is clearly not the case, showing that the factorization (19) does not hold in general.

It seems that the definition of dimension-wise independence is not an absolute one. One can either drop it altogether, or if need arises, change it to the following:

Definition 2 *An MIRT model given by an IRHS is* dimension-wise independent *if there exists a coordinatization of abilities so that the functional representation of the model $f_{a,b}(\theta)$ can be written as a product of factors*

$$f_{a,b}(\theta) = \prod_{d=1}^{D} h(a_d, b_d, \theta'_d). \tag{23}$$

The specialty of this property comes from the fact that for a general IRHS it is very rare that the functional representation can be factored so that one may consider it dimension-wise independent. This interpretation was used throughout the paper, when the Whitley model was called dimension-wise independent.

4 Conclusion

A coordinate-free definition of MIRT has been put forward in the paper. The main argument is that in a coordinate-free setup it is easier to distinguish genuine MIRT objects from potential artifacts. These artifacts can be notions and relationships, either of which should not be considered integral parts of the model since key features apparent in one coordinate system could vanish in another. This paper showed that it is possible to provide a full classification of monotonic models solely based on general, coordinate-free considerations.

It is very important that the reader does not mistake the promotion of the coordinate-free description as an argument for a completely coordinate-free handling of the entirety of MIRT. In fact, meaningful MIRT practice cannot exist without a choice of coordinates. In addition to this, every discussion of MIRT features can be fully carried out using $\mathbb{R}^D$ as the main model space for abilities. Should such a path be chosen, however, one has to be careful to meticulously maintain the coordinate-system invariance of the theory every step of the way. The contribution of this paper is an introduction of a framework to ease this burden by keeping the presentation *absolute* (without choosing any coordinates) for as long as possible. The paper shows that one may be able to formulate general statements and reach valuable insights before switching to *relative* mode by an introduction of a particular basis. It is likely that someone may observe the relevance of a notion while using a particular coordinate system and may want to establish whether the notion is invariant by trying to create a definition in the absolute framework presented here.

It is noteworthy that the necessity of the existence of a coordinate-free representation of our physical world led Einstein to formulate both the special and the general theories of relativity (1905, 1916). The fundamental dogma in relativity theory is that the events of the physical world take place without there being aware of any coordinate system. Therefore, any faithful description should be invariant of the change of coordinate system. Better yet, a description of the physical world is sought that bypasses the use of coordinates altogether.

A reader interested in the successes of coordinate-free description of the physical world may also find the books by Matalocsi (1986, 1993) useful.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: MIT Press.

Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik*, *17*, 891–921.

Einstein, A. (1916). Grundlagen der allgemeinene relativitätstheorie (The foundation of the general theory of relativity). *Annalen der Physik*, *49*(4), 284–339.

Halmos, P. R. (1974). *Finite-dimensional vector spaces* (2nd ed.). New York: Springer-Verlag.

Matolcsi, T. (1986). *A concept of mathematical physics: Models in mechanics.* Budapest, Hungary: Akadémiai Kiadó.

Matolcsi, T. (1993). *Spacetime without reference frames.* Budapest, Hungary: Akadémiai Kiadó.

McKinley, R. L., & Reckase, M. D. (1982). *The use of the general Rasch model with multidimensional item response data* (Research Rep. ONR No. 82-1). Iowa City, IA: ACT.

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271–286). New York: Springer-Verlag.

Warner, F. W. (1971). *Foundations of differentiable manifolds and lie groups.* Glenview, Illinois: Scott, Foresman and Company.

Whitely, S. E. (1980). *Measuring aptitude processes with multicomponent latent trait models* (Technical Rep. No. NIE-80-5). Lawrence: University of Kansas.

Zhang, J., & Stout, W. F. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, *64*, 129–152.